

GUIDELINES FOR CORE MEASURES

CENTER FOR SUBSTANCE ABUSE PREVENTION

GUIDELINES FOR CORE MEASURES

TABLE OF CONTENTS

INTRODUCTION	1
OVERVIEW	2
AT A GLANCE	3
WHY MEASURE PROGRAM IMPACT?	4
PRINCIPLES OF MEASUREMENT	5
SELECTING CORE MEASURES FOR ADMINISTRATION	5
SUBSTITUTIONS AND SUPPLEMENTS	6
REGULATORY AND LEGAL MATTERS	6
MEASUREMENT IS A SCIENCE	7
SCHEDULING MEASUREMENT SESSIONS AND DATA COLLECTIONS.....	8
CODING AND ANALYZING MEASUREMENT DATA	8
GLOSSARY OF MEASUREMENT AND EVALUATION TERMS	10

INTRODUCTION

When we visit the doctor's office for an illness, we often leave with a prescription and a set of instructions for using the medicine and any other regimen the doctor orders. If we don't use the medicine correctly, or if we fail to follow the doctor's regimen, we may not recover as quickly as if we had followed the instructions.

In the foregoing example, the prescription is a tool the doctor uses to fix our ailing body. Incorrect use of the tool will not fix the body as correct use will. Although the tool may be powerful, it alone may not do good and may even harm us if applied in the wrong manner. The same is true with other tools. They must be used as intended. Those intentions for use, in turn, are usually conveyed through a set of instructions or guidelines.

Measurement instruments, such as the core measures you have received for evaluating substance abuse prevention programs, can also be regarded as tools. Indeed, researchers who construct, apply, and test measurement instruments often refer to measures as tools. Just as a prescription is a tool aimed to bring good health, a measure is a tool designed to effect good evaluations. But like good medicine, measurement tools cannot be employed without careful instructions. Careless use of these tools may invite poor results.

No matter how reliable and valid, a measure is only as valuable as the hands that administer it. Much damage has occurred over the years by measures being incorrectly used. We need look no further than inappropriate applications of IQ tests to see that damage. Historically, the results of IQ tests were the basis for making monumental decisions over the lives of children and adults. When those tests lacked sensitivity for the population to whom they were applied, their results were worse than erroneous: they were harmful when they misdirected such activities as school tracking, college admissions, and job placement.

Though some would dispute the wisdom of giving any IQ test as a measure of intelligence and potential, the misapplication of these tests and their findings probably account for more damage than the tests themselves. Here again, we witness the importance of proper instructions for using tools. This time, our example concerned the tools of measurement. So you will correctly use the tools of core measurement that we are recommending for substance abuse prevention programs, these GUIDELINES provide you with correct and detailed instructions in the measures' use.

OVERVIEW

This initial set of core measures for evaluating the outcomes of SIG-funded substance abuse prevention programs have undergone a serious process of development. Task forces and working subgroups have been convened and charged with compiling the best-available measures for known facets of substance use knowledge, attitudes, and behaviors. Task force and subgroup members have, within the limits of time and resources, combed the scholarly and practice literature and reviewed work in the field.

The resulting compilation of core measures was assayed by CSAP to ensure that the final list of selected measurement instruments met criteria relevant to the SIGs and pertinent to the domains and constructs targeted by local programs. The core measures now part of this package represent that final distillation of state-of-the-art of substance abuse measurement instruments.

Besides reflecting the latest and updated information on substance abuse outcome measurement, these core measures are proven, tested, and user-friendly. Each measure has undergone rigorous study into its psychometric properties. Each measure has been employed successfully in prior evaluations similar to the ones you are undertaking.

The following sections of these GUIDELINES will describe in detail how you can use the measures to determine the success of your own program. For program planners, implementers, and evaluators who want the essence of the GUIDELINES, an initial section gives key information on use of the core measures. This “At a Glance” section serves as the ultimately summarized guide. Most of you will prefer to read all of the GUIDELINES. For you, the 10 points covered in “At a Glance” will act as reminders of what you read.

Following the “At a Glance” section is the rationale for measurement in outcome evaluations of substance abuse prevention programs. The GUIDELINES then review key principles of measurement. A glossary of measurement and evaluation terms is provided to define concepts used in this document and employed elsewhere by prevention program evaluators. Finally, a feedback sheet invites comments on the GUIDELINES so we can improve their responsiveness in subsequent iterations.

AT A GLANCE

1. Do not administer all core measures to all program participants. Instead, give only those measures that relate to constructs that your program targets for change.
2. Use additional measures if you wish when the measures address constructs and domains relevant to your program.
3. You do not need to substitute core measures for measures you are already administering. Instead, you may use the core measures to supplement your existing measures. When core measures and existing measures assess the same construct or behavior, you are the best judge of whether to substitute or supplement. The core measures are useful to assess program effectiveness with those populations upon which they were previously tested. If the measure hasn't been tested for your program's target population, we still would like to encourage its use as this will contribute substantially to our knowledge of its appropriateness for particular populations. However, to assess program effectiveness, you should use a tool that has already been tested with your targeted population.
4. Measuring substance abuse knowledge, attitudes, and behaviors is a science. If you cannot follow carefully the scientific principles of measurement, do not administer the core measures. Bad measurement is worse than no measurement.
5. Do not combine or alter measures unless you know how those changes affect the original measure. Only expert evaluators should change a measurement instrument.
6. Handling, analyzing, and reporting outcome data are as important as collecting data. Each step in the process must be followed with the same level of care.
7. In most instances, when collecting data from minors, you will need prior and signed permission from parents or legal guardians and the tacit approval, or assent, or youths themselves. Institutional Review Boards (IRGs) regulate such data collection activities.
8. Many states have laws that require the reporting of information that minors provide in the course of data collection. You should check your own state's laws prior to data collection.
9. CSAP would like copies of all measures being used to evaluate SIG-funded programs.
10. Completing the feedback sheet at the end of these GUIDELINES will help CSAP design clearer and better GUIDELINES in the future.

An excellent compendium of measures for prevention programs is available from CSAP and should be consulted for in-depth coverage of measurement principles, procedures, issues, and a panoply of other important measurement considerations. The compendium is entitled:

Measurements in Prevention

A Manual on Selecting and Using Instruments to Evaluate Prevention Programs

The book is listed as CSAP Technical Report no. 8, and its authors are Karol L. Kumpfer, Gail H. Shur, James G. Ross, Kate K. Bunnell, John J. Librett, and Allison R. Millward. It was published in 1993.

WHY MEASURE PROGRAM IMPACT?

As professionals invested in the programs we develop and manage, we cannot help but believe that our efforts are having a beneficial impact on members of our consumer groups. In the case of substance abuse prevention programs, we must have confidence that any intervention we offer young people will help them stay away from smoking, drinking, and drug use. If we lack faith in the wisdom of our prevention programs, we would probably not continue to deliver them. Human that we are, however, we are vulnerable to the same biases that affect other professionals. That is, we tend to be overly optimistic about the value of our works.

Though we may think that participants are benefiting from well-intentioned efforts, they may in fact remain unchanged after receiving a prevention program. Although rare, prevention programs could actually disserve participants by giving them erroneous information or by employing untested strategies that increase rather than decrease substance use and experimentation. Moreover, those who fund prevention programs may rightly want objective, unbiased findings on whether interventions work, with whom, under which conditions, and at what cost.

For these reasons, we who are involved in prevention increasingly find ourselves needing to evaluate our efforts. Evaluation pressures are coming to bear on programs that heretofore were regarded as sacrosanct and inherently good and valuable for a community. With competing demands for scarce public and private funds for human services programs, those at the community level are evermore faced with the necessities of measuring what we do.

Measurement of program outcomes has added benefits for programs that receive SIG funding. When government agencies begin to appreciate the payoffs of substance abuse prevention efforts, they may be more inclined to invest additional resources in these activities. Results of good evaluations can provide compelling evidence for new and increased initiatives in prevention. Quite simply, measuring program outcomes has many reasons to justify its inclusion in any prevention effort.

Consider these benefits of program impact data:

- Program developers will have first-hand, objective evidence of their successes and failures to appropriately justify their prevention efforts or to change their modus operandi toward making programs effective.
- Funders will receive concrete findings on their return on investment.
- Tax payers will witness the tangible human capital advantages of such intangibles as substance abuse prevention programs.
- Most important, America's citizens will benefit when prevention programs are subjected to the cold scrutiny of scientific evaluation and either expanded when effective, or modified when ineffective.

PRINCIPLES OF MEASUREMENT

In this section, we cover regulatory, scientific, and logistical information related to collecting data for outcome assessment of substance abuse prevention programs. Hardly exhaustive, the list of principles in this section nonetheless provide a thorough review of most issues program evaluators encounter in outcome measurements with children and youth relative to substance abuse prevention.

SELECTING CORE MEASURES FOR ADMINISTRATION

Neither CSAP nor anyone else intends for you to administer all the core measures or even most of the core measures to your program participants. Rather, the list of core measures is provided as a menu from which you select only those measures that relate to constructs your prevention program has targeted for change. For example, if you are targeting substance use as an outcome and believe that your program will prove successful to the extent that you reduce youths' tobacco, alcohol, and other drug use, then you should measure that use with the scales recommended by the CMI.

Continuing that example, if your measurement battery does not currently include substance use items, you will want to adopt the respective core measure. If your program does not target substance use and instead views intermediary variables as important – knowledge and attitudinal change, as examples – then you would not administer core measures of substance use, but you would administer core measures for the knowledge and attitudes constructs you are targeting. Also, please note whether the recommended core measure has been tested on your targeted population. If it has not, you should use a tested measure appropriate for your population. However, we would like to encourage you to also use the core measure so that we could find out whether it could be recommended for use in the future with your targeted population.

How many core measures are enough? Use additional measures if you wish and when the measures address constructs and domains relevant to your program. Many of you will want to add a variety of core measures to enhance your overall data collection efforts. This can be advantageous if it elicits important data about program impact. But more is not better if you overtax your respondents, do not collect reliable information, or ask about constructs and domains not pertinent to your program.

Example:

XYZ Youth Services Inc. is offering an after-school drug abuse prevention program for its client population of 10- to 14-year olds.

- Because the program is interested in youths' actual substance use, its outcome evaluation will focus on the domain of ATOD and will administer the Lifetime Use scale from the Monitoring the Future Study.
 - XYZ's program also focuses on the individual and peer domain and, because the intervention aims to increase youths' self-esteem, its outcome evaluation will employ Rosenberg's Self-Esteem Scale.
 - Finally, the program's emphasis on the family warrants inclusion of family conflict items from the SDRG Student Survey. Though the XYZ evaluation team could have administered many more core measures, because they already use measures of their own – which they share with others in their community and state and with CSAP – these three additional measures covering as many domains will suffice.
-

SUBSTITUTIONS AND SUPPLEMENTS

Core measures are not intended as substitutions for scientifically sound outcome measures you are already using. You do not have to substitute core measures for proven measures currently in your administration battery. Instead, use core measures to supplement your existing measures. When core measures and existing measures assess the same construct or behavior, you are the best judge of whether to substitute or supplement.

In other instances, measures currently employed by program providers will lack questions available in core measures. Then, providers should include the core measure, *as long as the construct measured is one that the program seeks to modify*. In still other instances, providers will have developed measures that address variables not covered in the core measures or covered differently in the core measures. If this circumstance applies to your program, CSAP is eager to receive copies of your questionnaires. You may have identified a way of assessing a variable that the core measures overlook.

REGULATORY AND LEGAL MATTERS

Data collection for research or program evaluation purposes is increasingly regulated. For historical reasons and simply because it makes good sense, you must be particularly mindful of the rights of minors and their parents and guardians when soliciting any information from children and youth under 18 years of age. Usually, data collection activities must be sanctioned by an Institutional Review Board. Though a full procedural manual on IRB operations is well beyond the scope of these GUIDELINES, you should be familiar with the need for IRB approval, if it is applicable to your data collection.

The best resource for learning about the IRB process is your local evaluator. While most small community agencies do not have their own IRB, you can still gain access to the services of one through the auspices of your local evaluator, sub-recipient, or state agency. However you interact with an IRB, it will provide guidelines for your data collection, including the manner in which you must gain the informed consent of your data collection participants.

Usually, parents or legal guardians, must give permission for data collections involving their children or wards. That permission is best given in writing. Minors may not be asked for written permission *per se*, but they should give their assent prior to any data collection. These and other issues regarding informed consent are discussed in detail on the website of the respective Federal agency that governs human subjects review procedures. That agency is the Office for Protection from Research Risks. Its website address is: www.nih.gov/grants/oprr.htm

Asking people about their personal drug use may also expose them and you to the applicability of other Federal and state laws. For example, if someone tells you that they are using or have used a controlled substance, they are essentially revealing that they are breaking or have broken the law. In certain circumstances, that admission could be used as evidence in a legal proceeding. Because of such exigencies, many drug abuse surveys are completed anonymously or under strict confidentiality requirements that protect the data from subpoena. As with human subjects protection considerations, matters related to legal issues around data collection are best handled by an IRG or through an appropriate local, sub-recipient, or SIG program state agency. Although Federal laws also govern drug use activity, the best initial source for information about data collection sanctions is the body from which you derive your funding.

Finally, state laws often require that certain other data reported by minors be forwarded to appropriate agencies for action. For instance, if a child reports abuse or neglect in the course of your questioning the child about prevention program effects, you must in turn report that information to the relevant child protective services office. Here again, you should consult with your local funding agency, sub-recipient, or SIG program contact for detailed instructions on such reporting, including your obligations and those of your evaluators.

MEASUREMENT IS A SCIENCE

To their detriment, social scientists often compare themselves to medical scientists and those who study in the fields of natural sciences. Not infrequently, these comparisons result in social science being called “soft”, whereas physical sciences are termed “hard.” In fact, the sciences have much in common. Like their physical science brethren, social scientists follow a rigorous discipline in carrying out their research. A large part of that discipline applies to measurement and data collection.

Just as done in the physical sciences, social science relies on conventional standards and rules of measurement. To adhere to those standards and to follow those rules means that information gathered in the course of research has integrity and is respected. But to violate the principles of measurement and data collection is to place at jeopardy the very findings that we seek in evaluating prevention programs. In short, data can only be trusted if they were gathered properly.

Here are 12 conventional rules of measurement that cannot be violated:

- a) Use a measurement instrument (such as those included in the core measures) in its entirety or not at all. To only use part of a measure is to use the measure incorrectly. Findings from partial administration of a measure cannot be compared with other administrations of that measure.
- b) Retain the wording and order of items in a measure as they appear in the original. Changing the items can destroy the measure’s integrity.
- c) Administer measures in a quiet, confidential, non-distracting environment. A measurement setting must be free of opportunities for participants to observe one another’s answers to measurement items. Likewise, youths should not believe that adults can see their responses as they are recording them. The measurement session should be treated like an academic testing situation.
- d) Honestly tell respondents whether their responses are anonymous or confidential and whether and with whom their data will be shared. Anonymous measures are those with no identifiable information. Confidential data are those not shared with certain individuals. Respondents must know who will see their data and who will not see their data. Do not confuse the terms anonymous and confidential. Participants will know the difference and, should you mix up the terms, respondents will resent your misrepresenting the nature of security over their responses.
- e) Assure participants that program outcome measures have no right or wrong answers. The only wrong answer is one that fails to accurately portray what the respondent intended. Otherwise, every answer is right.
- f) Ensure that respondents complete all items on a questionnaire before they leave the measurement session. Scan quickly questionnaires as respondents complete them to provide this assurance.
- g) Do not comment on any questionnaire responses in the presence of any respondent– regardless of whether they were involved in the measurement process. Never provide feedback on their own or anyone else’s questionnaire responses.
- h) If a respondent has difficulty understanding an item, other respondents may have the same difficulty. Take pains to explain the meaning to the item to all respondents. Make a clarifying announcement prior to administration of the questionnaires such as the following: “Some people find the wording of Question 3 confusing. What it’s asking is, ‘How many individual marijuana cigarettes or joints have you smoked in the last week.’”
- i) When respondents have underdeveloped reading skills, read aloud each measurement item and allow respondents to *confidentially* record their answer.

- j) Collect and seal in an envelope all questionnaires before respondents and you leave the measurement session.
- k) Secure and guard completed measures as you would an envelope full of \$100 bills. Indeed, outcome data demonstrating program effects are worth many times that amount. Once lost, most data are irreplaceable. What is more, data that fall into the wrong hands can bring embarrassment at best, lawsuits at worst. Keep all data locked and under close scrutiny.
- l) Once entered and coded, data should be continue to be secured. Original questionnaires from which those data issued can be destroyed, but only in a manner that will ensure their permanent confidentiality. Shredding original and personally identifiable measurement responses is prudent and conservative.

By following these 12 rules and exercising common sense and good judgement, you will collect measurement data of the highest quality and that show respect for the rights of and protections due your program participants and their families.

SCHEDULING MEASUREMENT SESSIONS AND DATA COLLECTIONS

To determine whether a prevention program had an impact on its participants, measurement instruments need to make comparisons. Those comparisons may be between participants who received a program and those who did not. Or, comparisons may include participants who received two types of programs. Various other comparisons are options as well, depending on the goals and design of the evaluation.

Regardless of evaluation design and strategy, however, any determination of programmatic impact must include a time comparison. At least two time periods are always necessary: before the program begins and after it ends. Respectively, these time periods are called pretest and posttest measurement occasions. Since prevention programs aim to keep something from happening – to keep participants from using substances, in our instance – additional time comparisons are helpful as participants develop, mature, and face increasing pressures and opportunities to smoke, drink, and use drugs.

Program providers must therefore schedule their measurements in line with the foregoing parameters. Whenever multiple measurements are scheduled, the same questionnaires and procedures must be employed at each occasion. Otherwise, data from one measurement occasion are not comparable with those obtained at another occasion.

CODING AND ANALYZING MEASUREMENT DATA

A singular advantage of administering core measures is that each instrument has a standard coding protocol. Moreover, each core measure has been applied extensively in prior prevention evaluations and thus has precedents for the preferred manner in which data issuing from it should be analyzed. Instructions on that coding and those analyses are found in literature that support each core measure.

The advent of user-friendly computer programs has made data entry, coding, and analysis easier and faster than ever before. As a result, these formerly complex research tasks can appear deceptively simple. In fact, they remain complex. Each task should only be undertaken by evaluators who have training and experience in this area.

Rather than becoming versed in the language and techniques of data analysis, program-level professionals can better devote their attention to formulating questions to which they want answers. The following are some of the kinds of evaluation questions that may put to the data and answered empirically, depending, of course on the nature of the data collection procedures and evaluation design.

- To what degree did participants change their knowledge, attitudes, and/or behaviors during the course of the program? Were these changes sustained longitudinally?

- Does a causal relationship exist between the prevention program and participant outcomes?
- Did particular participants (grouped, for example, by gender, age, ethnicity, and geography) benefit disproportionately from the program?
- Is there a relationship between the amount of intervention received and prevention program outcomes?
- Based upon measurement of different constructs and domains, which programmatic elements appear to account for the greatest changes? For the smallest changes?
- How generalizable are evaluation outcome findings to the larger populations from which participants were drawn?
- Can we determine the cost-effectiveness or cost-benefits of the program, based on outcome data and what we know about the expenses consumed by the program?
- What level of confidence can we place in evaluation findings?
- Can we identify the greatest strengths and limitations of the evaluation, based on outcome data?

By properly assuming their roles in program design and implementation, community level staff can put their energies in determining what is best for the consumers of prevention services and delegate to evaluation staff duties of data entry, coding, and analysis.

GLOSSARY OF MEASUREMENT AND EVALUATION TERMS

Hardly exhaustive, this list of definitions covers many terms used in measurement and evaluation. Some of the terms are used in the present GUIDELINES; other terms will arise in conversations with evaluators, data collectors, and data analysts.

Causal Relationship: Theoretical notion that change in one variable forces, produces, or brings about a change in another. Requires empirical support through a controlled experiment.

Content Analysis: Process of studying and tracking what has been written and discussed, then translating this qualitative material into quantitative form through some type of counting approach that involves coding and classifying of specific messages.

Correlation: Association or relationship between two variables.

Correlation Coefficient: Measure of association (symbolized as r) that describes the direction and strength of a linear relationship between two variables, measured at the interval or ratio level (e.g. Pearson's Correlation Coefficient).

Cross-Sectional Study: Study based on observations representing a single point in time.

Ethnographic Research: Relies on the tools and techniques of cultural anthropologists and sociologists to obtain a better understanding of how individuals and groups function in their natural settings. Usually, this type of research is carried out by a team of impartial, trained researchers who immerse themselves into the daily routine of a neighborhood or community, using a mix of observation, participation, and role-playing techniques, in an effort to try to assess what is really happening from a cultural perspective.

Evaluation Research: Determines the relative effectiveness of a particular program or strategy, measuring outputs and outcomes against a predetermined set of objectives.

Experiment: Controlled arrangement and manipulation of conditions to systematically observe specific occurrences, with the intention of defining those criteria that might possibly be affecting those occurrences. An experimental, or quasi-experimental, research design usually involves two groups – an experimental group exposed to given criteria, and a control group, not exposed. Comparisons are then made to determine what effect, if any, exposures to the criteria have had on those in the experimental group.

Factor Analysis: Algebraic procedure that seeks to group or combine items or variables in a questionnaire based on how they naturally relate to each other as general descriptors or factors.

Focus Group: Exploratory technique in which a group of somewhere between 8 and 12 individuals – under the guidance of a trained moderator – are encouraged, as a group, to discuss freely any and all of their feelings, concerns, problems and frustrations relating to specific topics under discussion. Focus groups are ideal for brainstorming, idea-gathering, and concept testing.

Hypothesis: Expectation about the nature of things derived from theory.

Hypothesis Testing: Determining whether expectations that a hypothesis represents are, indeed, true.

Incidence: Frequency with which a condition or event occurs within a given time and population.

Likert Scale: Developed by Rensis Likert, composite measure in which respondents are asked to choose from an ordered series of five responses to indicate their reactions to a sequence of statements (e.g., strongly agree ... somewhat agree ... neither agree nor disagree ... somewhat disagree... strongly disagree).

Longitudinal Study: Research design involving the collection of data at different points in time.

Mean: Measure of central tendency which is the arithmetic average of the scores.

Median: Measure of central tendency indicating the midpoint in a series of scores, the point above and below which 50% of the values fall.

Mode: Measure of central tendency which is the most frequently occurring, the most typical, value in a series.

Multivariate Analysis: Examination of the relationship among three or more variables.

Panel Study: 1) Type of longitudinal study in which the same individuals are interviewed more than once over a period of time to investigate the processes of response change, usually in reference to the same topic or issue. 2) Also, type of study in which a group of individuals are deliberately recruited because of their special demographic characteristics, to be interviewed more than once over a period of time on different topics or subjects.

Probability Sample: Process of random selection, in which each unit in a population has an equal chance of being included in the sample.

Qualitative Research: Usually refers to studies that are somewhat subjective, but nevertheless in-depth, using a probing, open-end, free-response format.

Quantitative Research: Usually refers to studies that are highly objective and projectable, using closed-end, forced-choice questionnaires. These studies tend to rely heavily on statistics and numerical measures.

Range: Measure of variability that is computed by subtracting the lowest score in a distribution from the highest score.

Regression Analysis: Statistical technique for studying relationships among variables, measured at the interval or ratio level.

Reliability: The extent to which the results would be consistent, or replicable, if the research were conducted a number of times.

Secondary Analysis: Technique for extracting from previously conducted studies new knowledge on topics other than those which were the focus of the original studies. Usually involves systematic re-analysis of existing data.

Standard Deviation: Index of variability of a distribution. Range from the mean within which approximately plus or minus 34% of the cases fall, provided the values are distributed in a normal curve.

Statistical Significance: Refers to the unlikeliness that relationships observed in a sample could be attributed to sampling error alone.

Survey: Systematic collection of data that uses a questionnaire and a recognized sampling method. Surveys are conducted face-to-face (in-person), by telephone, and are self-administered (usually distributed by mail, e-mail, or fax.)

Univariate Analysis: Examination of only one variable at a time.

Validity: Extent to which a research project measures what it is intended, or purports, to measure.

Variance: Measure of the extent to which individual scores in a set differ from each other. Sum of the squared deviations from the mean divided by the frequencies.